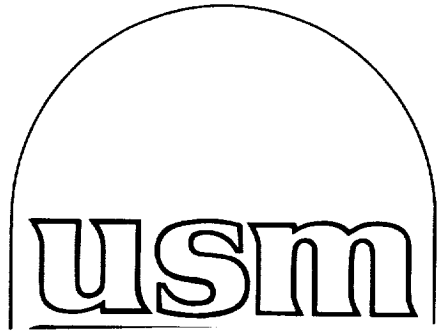
NASA-CR-195750

College of Science and Technology



(NASA-CR-195750) CONTINUATION OF THE DEVELOPMENT OF SOFTWARE TO BE USED IN SUPPORT OF THE NASA TECHNOLOGY UTILIZATION PROGRAM Annual Report, 1 Jul. 1985 - 30 Jun. 1986 (University of Southern Mississippi) 33 p

N94-71788

Unclas

Z9/61 0002874

SCIENCE AND TECHNOLOGY

CONTINUATION OF THE DEVELOPMENT OF SOFTWARE TO BE USED IN SUPPORT OF THE NASA TECHNOLOGY UTILIZATION PROGRAM

G. David Huffman

Annual Report
July 1, 1985 - June 30, 1986

CONTINUATION OF THE DEVELOPMENT OF SOFTWARE TO BE USED IN SUPPORT OF THE NASA TECHNOLOGY UTILIZATION PROGRAM

ANNUAL REPORT:

July 1, 1985 - June 30, 1986

G. David Huffman, Ph.D.

July, 1986

College of Science and Technology
University of Southern Mississippi
Southern Station, Box 5165
Hattiesburg, Mississippi 39406

EXECUTIVE SUMMARY

Technology transfer, research and development and engineering projects frequently require in-depth literature reviews. These reviews are carried out using computerized, bibliographic databases. The review and/or searching process involves keywords selected from database thesauri. The search strategy is formulated to provide both breadth and depth of coverage and yields both relevant and non-relevant citations. Experience indicates that about 10-20% of the citations are relevant. As a consequence, significant amounts of time are required to eliminate the non-relevant citations. This report describes statistically-based, lexical association methods which can be employed to determine citation relevance. In particular, the searcher selects relevant terms from citation-derived indexes and this information along with lexical statistics is used to determine citation relevance. Preliminary results are encouraging with the techniques providing an effective concentration of relevant citations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

Variable	Description
B_k	Number of documents in which the k-th term occurs in J citations
$b_{j,k}$	Binary occurrence of the k-th term in the j-th citation
D_k	Discrimination factor for the k-th term
F_k	Number of occurrences of the k-th term in J citations
N_k	Noise of the k-th term in J citations
$n_{j,k}$	Number of occurrences of the k-th term in the j-th citation
$\hat{n}_{j,k}$	Normalized number of occurrences of the k-th term in the j-th citation
R_j	Ranking of the j-th citation
r_k	Relevance of the k-th term
S_k	Signal of the k-th term in J citations
W_k	Weighing factor for the k-th term

SUBSCRIPTS

Variable	Description
J	Total number of citations
j	Citation
K	Total number of terms
k	Term

SUPERSCRIPTS

Variable	Description
f	Frequency
s	Signal

1. SEMI-AUTOMATIC DETERMINATION OF CITATION RELEVANCY

1.1 Introduction

The automation of the citation review process has been carried out in conjunction with an in-depth review of technology transfer processes. The review and automation have been directed toward techniques and mechanisms used by NASA Industrial Application Centers. Particular emphasis has been placed on the industrial applications study [1]. This mode of technology transfer can be defined as the location of existing technology which addresses a specific need and its reformulation into a form which is usable and understandable to the recipient. This normally implies the secondary utilization or application of technology for purposes other than those for which the technology was originally intended or created. The process usually involves the crossing of disciplinary or mission barriers.

The actual processes involved in an industrial applications study are shown in flowchart format in Figure 1. The study is generally carried out by an engineer or scientist. The process consists of: the selection of databases, formulation of search strategies, conduct of on-line searches, review and ranking of citations for relevancy, ordering and analysis of publications, contact and consultation with experts, preparation of a final report, transmittal of the report to the client, and the carrying out of a benefits analysis.

The over-all process is labor intensive; however, a few of the elements are amenable to automation. In particular, pre-processors for database access [2], [3], [4], and [5] and post-processors for citation analysis are available. This report will discuss one aspect of post-processing--the semi-automatic determination of abstract relevancy.

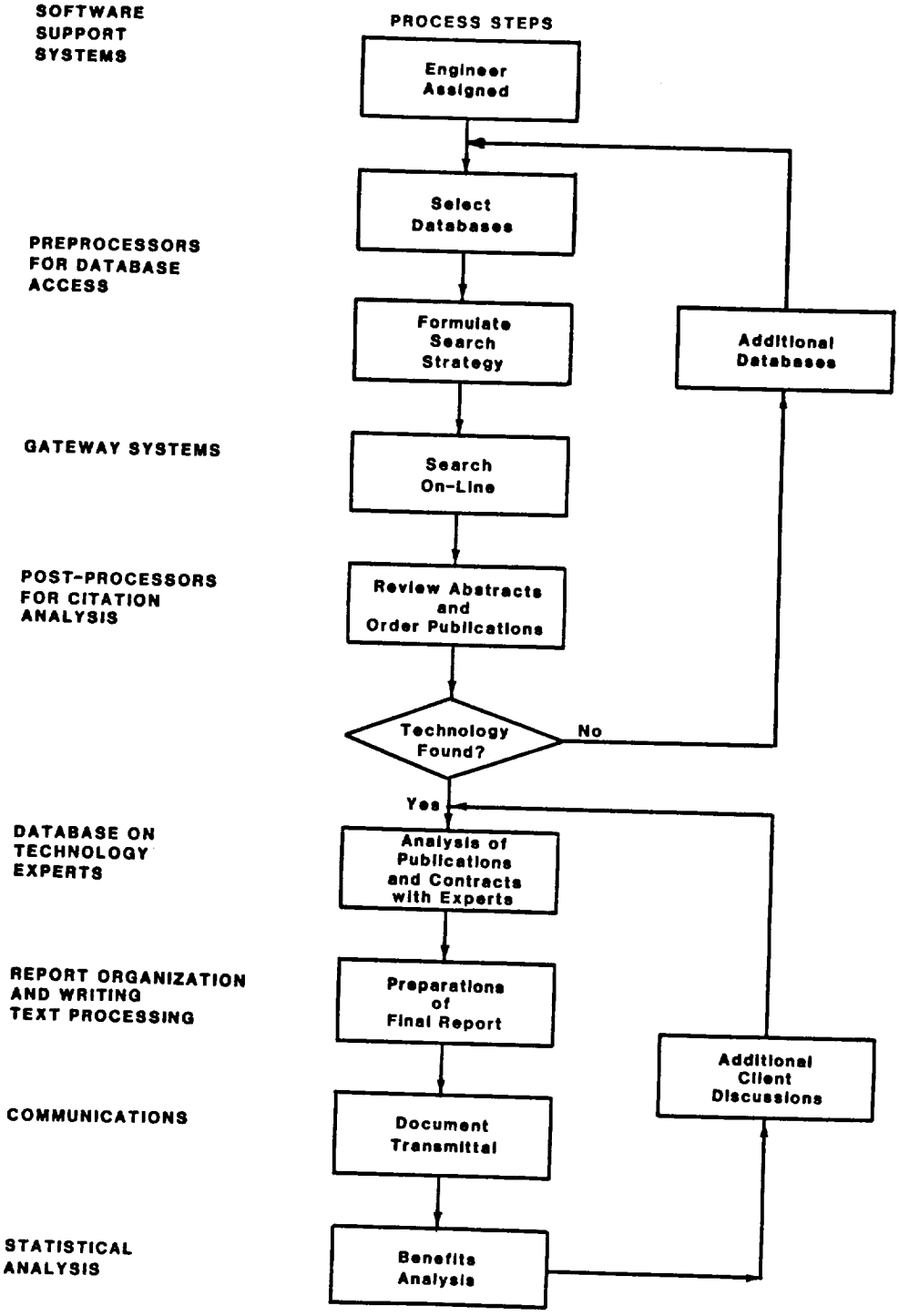


Figure 1. Steps Carried Out in an Industrial Applications Study

1.2 Over-all Approach

The method used to evaluate abstract relevancy combines three elements: a statistical and/or lexical analysis of the abstracts to determine a characteristic word index, a user determination of word relevance and the combination of user input and statistical analysis to rank the citations for relevance. The method and some preliminary results will be discussed in the following sections.

1.3 The Mathematical Model

As noted above, the relevancy determination combines lexical association methods [6], [7], [8], [9], and [10] with user input. The lexical and/or statistical techniques are used to generate an index of characteristic words. These are not selected from the controlled vocabulary keywords of the online database system but are generated with the lexical association algorithms from the post-search collection of citations. The user selects relevant words from this list. This information along with the statistical metrics is used to rank the citations for relevance.

The statistical methods used to generate the characteristic word indices are based on the word occurrence matrix. This is shown in Figure 2 and defines the number of occurrences of a given term in a specified citation, e.g., $n_{2,3} = 6$ implies that the third term appears 6 times in the second abstract. Note that terms lacking a report specific meaning, e.g., a, an, the, it, etc., are neglected by employing a stop list.

TERM		ABSTRACT				
k	j:	1	2	3	. . .	J
1		$n_{1,1}$	$n_{2,1}$	$n_{3,1}$	. . .	$n_{J,1}$
2		$n_{1,2}$	$n_{2,2}$	$n_{3,2}$	. . .	$n_{J,2}$
3		$n_{1,3}$	$n_{2,3}$	$n_{3,3}$	. . .	$n_{J,3}$
.		.				
.		.				
.		.				
K		$n_{1,K}$	$n_{2,K}$	$n_{3,K}$	. . .	$n_{J,K}$

Figure 2. Occurrence Matrix, $n_{j,k}$

While all the various statistical parameters could be based on the occurrence matrix, it was found that a normalized form was more effective. The normalized form uses the median abstract size and is defined as

$$\hat{n}_{j,k} = \text{INTG}\left\{ \frac{n_{j,k} \sum_{k=1}^{K} n_{j,k}}{\sum_{j=1}^{J} \sum_{k=1}^{K} n_{j,k}/J} \right\} \qquad (1)$$

where $\hat{n}_{j,k}$ is rounded to the nearest integer except that $\hat{n}_{j,k}$ is never rounded to 0 for $n_{j,k}$ greater than 0. $\hat{n}_{j,k}$ is effectively a frequency of occurrence with the advantage of integer arithmetic. This is a significant factor since the software system is configured for operation on a fixed disk, microcomputer system.

A number of parameters can be derived from the $\hat{n}_{j,k}$ term. In particular, the frequency of the term k in the collection can be defined as

$$F_k = \sum_{j=1}^{J} \hat{n}_{j,k} \qquad (2)$$

Even though F_k is based on a normalized occurrence matrix, it can be biased by a few large values of $\hat{n}_{j,k}$. The document occurrence frequency, i.e.,

$$b_{j,k} = \begin{cases} 1 & \hat{n}_{j,k} > 0 \\ 0 & \text{Otherwise} \end{cases} \qquad (3)$$

$$B_k = \sum_{j=1}^{J} b_{j,k} \qquad (4)$$

is not subject to this influence. Both F_k and B_k are measures of the importance of the term k with the former favoring total occurrences and the latter occurrences in multiple documents.

The signal-noise ratio of communication theory [11] can also be applied to the citation collection. The noise is defined as

$$N_k = 1.44 \left[\ln(F_k) - \frac{1}{F_k} \sum_{j=1}^{J} \hat{n}_{j,k} \ln(\hat{n}_{j,k}) \right] \qquad (5)$$

where

$$\begin{aligned} \hat{n}_{j,k} \ln(\hat{n}_{j,k}) &\longrightarrow 0 \\ \hat{n}_{j,k} &\longrightarrow 0 \end{aligned} \qquad (6)$$

by L´Hospital´s rule. The signal is defined as

$$S_k = 1.44 \ln(F_k) - N_k \qquad (7)$$

Note that the factor 1.44 results from the conversion of $\log_2 X$ to $\log_e X$, i.e., $\ln X$.

While both F_k and B_k are directly linked to $\hat{n}_{j,k}$, the relationship between N_k, S_k and $\hat{n}_{j,k}$ is less obvious. Some insight into the relationship between the variables can be gained by considering two specific cases. In the first case, let $\hat{n}_{j,k} = 1$, for all j and a specified k, i.e., the same term appears in all abstracts. Equations (2), (5) and (7) show that $F_k = J$, $N_k = 1.44 \ln(J)$ and $S_k = 1.44 \ln(J) - 1.44 \ln(J) = 0$. It thus follows

that the uniform occurrence of a term produces a background noise level but no signal.

In the second case, let $\hat{n}_{j,k} = 0$ for all abstracts save one with that value being F_k. Equations (2), (5) and (7) yield $F_k = F_k$, $N_k = 1.44\{\ln(F_k) - F_k\ln(F_k)/F_k\} = 0$ and $S_k = 1.44\ln(F_k)$. The multiple occurrence of a term in an individual abstract produces a non-zero signal with a 0 noise level.

The two cases represent the extremes. The relatively uniform presence of a term in many abstracts produces little signal and high noise. The infrequent appearance of a term will produce high signal levels with low noise. In a sense, F_k and S_k are complimentary in that F_k reflects numbers of occurrences while S_k depicts deviations from uniformity.

The weighing factors for the various terms which appear in the collections can be computed as

$$W_k^{(f)} = F_k D_k \qquad (8)$$

$$W_k^{(s)} = F_k D_k S_k \qquad (9)$$

where D_k is a discrimination factor [7] which is defined as

$$D_k = 1.44[\ln(J) - \ln(B_k) + 1] \qquad (10)$$

Equations (8) and (9) are used to define two sets (indices) of characteristic words. The number of words is arbitrary--all unique words are analyzed. Current versions of the program display the first 100 words, i.e., the words with the largest $W^{(f)}$ and $W^{(s)}$ values. The searcher assigns a relevance value,

r_k, to each of these words. r_k can take on any value desired with preliminary tests employing 0 for not relevant, 10 for moderately relevant and 20 for relevant.

The abstract and/or citation ranking equations combine both the W_k and r_k data. The equations employed are

$$R_j^{(f)} = \frac{1}{K} \sum_{k=1}^{K} \hat{n}_{j,k} D_k r_k^{(f)} \qquad (11)$$

$$R_j^{(s)} = \frac{1}{K} \sum_{k=1}^{K} \hat{n}_{j,k} D_k S_k r_k^{(s)} \qquad (12)$$

where R_j denotes the relative ranking of the J-th abstract. The superscripts f or s simply denote the metric employed.

1.4 Preliminary Results

The mathematical techniques of the previous section--along with a number of other features--have been implemented in a software system called SORT-AID [12], Figure 3. The system operates on a fixed-disk microcomputer with 512 Kbytes or more of memory. The software is now being evaluated by a number of NASA Industrial Application Centers with some preliminary results available. These will be discussed in the following paragraphs.

As noted in the previous section, the W_k equations are used to generate the indices of characteristic words. Two typical indexes are shown in Figures 4 and 5. Figure 4 lists the words generated with $w_k^{(f)}$ while Figure 5 shows those developed with $W_k^{(s)}$. The terms are listed by W_k values and these vary by a

factor of 10-30 for the first 100 words. The r_k values are user supplied. Note that these indexes are not derived from the controlled vocabulary of the database but from the citations themselves. Limited word stemming [10] is used to minimize duplication of terms which are essentially equivalent. Note that some terms are common to both lists while others are unique.

Equations (11) and (12) can be used to generate a series of R_k values. They are used to order the abstracts with the most relevant first followed by the least relevant. The actual determination of relevance is subjective and depends to some extent on the engineer evaluating the material. A number of experienced searchers have evaluated citations and determined relevance. The results to date are positive.

The relevance of the various citations is plotted in Figures 6, 7, 8, and 9. The objective of the process is to concentrate the relevant citations at the beginning of the collection. The data depicts this concentration by using abstract groups and relevant citations in the group/total number of relevant citations as the plotting parameters. Each of the four figures contains three separate curves--the ideal distribution, i.e., all the relevant citations grouped together in positions 1 to J_1 where J_1 denotes the total number of relevant citations, and distributions derived using the frequency, F_k, and signal-to-noise, S_k, approaches.

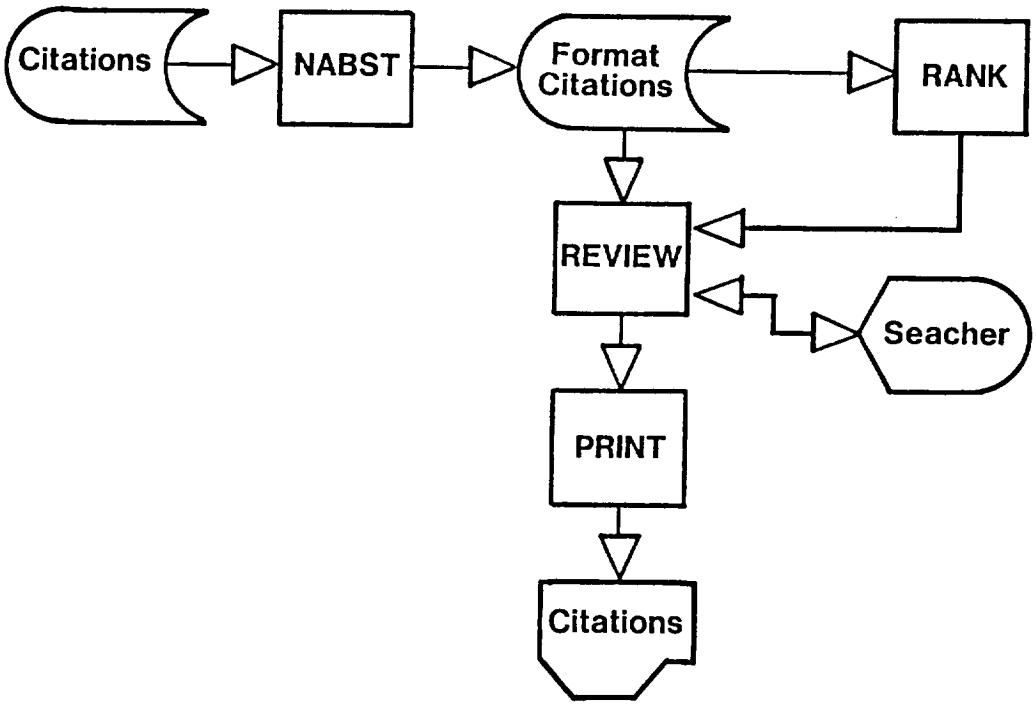


Figure 3. Schematic Diagram of the SORT-AID System

r_k	$Term_k$	W_k	D_k	r_k	$Term_k$	W_k	D_k
10	PUMP	1084	1.44	0	WATERJET	129	6.47
0	PRESSURE	647	1.99	0	MAXIMUM	127	4.09
10	HYDRAULIC	609	1.89	0	BAR	124	4.99
0	SEAL	569	4.15	0	AXIAL	120	4.47
0	SPEED	541	2.40	0	MOUNT	120	4.80
0	FLUID	490	1.77	0	DUE	118	4.55
0	CIRCUIT	450	3.75	0	ENGINEER	118	4.21
10	NOISE	449	4.27	0	MEAN	117	4.89
10	GEAR	432	3.40	0	JET	117	5.09
0	POWER	430	2.39	20	HYDROSTATIC	116	5.80
0	FLOW	390	2.69	0	LIFT	116	5.80
0	VALVE	319	3.59	0	LOAD	115	4.80
0	VANE	315	4.04	0	PROBLEM	115	4.80
0	RANGE	260	2.46	0	LEAKAGE	114	5.21
0	TURBINE	259	4.63	0	PSI	114	5.21
10	DISPLACEMENT	229	3.75	0	CONSTRUCTION	113	4.71
10	ROTARY	207	3.71	0	INTERNAL	112	4.89
0	OIL	206	4.21	0	ENERGY	112	5.09
0	DRIVE	205	3.67	0	SPECIFIC	111	4.63
0	VARIABLE	203	3.84	0	SIDE	109	5.47
0	FACE	200	4.89	0	PRODUCT	109	4.55
10	PISTON	195	4.15	0	OUTPUT	107	5.09
0	COUPL	189	4.99	0	PROPERTIE	107	5.09
0	SOLID	187	4.80	0	TOOL	106	5.63
0	PERFORMANCE	186	3.59	0	PROVIDE	106	4.63
0	CENTRIFUGAL	186	4.55	0	BEING	105	4.80
0	WATER	185	4.21	0	FIXED	103	4.71
0	LIQUID	173	4.34	0	SURFACE	101	5.09
0	MEASUREMENT	171	4.63	0	BEAR	101	5.99
0	STAGE	169	4.71	0	VERTICAL	101	5.99
0	OILHYDRAULIC	167	3.09	0	INVESTIGATION	101	5.34
0	BHRA	166	4.63	0	MODEL	101	5.34
0	IMPELLER	164	4.71	0	DISCUSS	101	4.40
20	TRANSMISSION	160	4.34	0	CLAIM	100	4.55
0	EFFICIENCY	157	4.04	0	PLANT	99	5.21
0	ENGINE	157	5.63	0	COST	98	4.71
0	DEVICE	156	4.89	0	AIRBORNE	98	5.80
0	TRANSPORT	156	5.21	0	WIDE	98	4.27
0	COMPONENT	155	3.80	0	SIZE	97	4.89
0	PROPULSION	155	6.47	0	HITACHI	97	6.47
0	CHARACTERISTIC	153	4.04	0	CONF	96	4.40
0	AIR	153	4.80	0	ADVANTAGE	96	5.09
10	VIBRATION	149	5.34	0	LENGTH	96	5.34
0	STANDARD	147	4.47	0	SUP	96	5.34
0	VORTEX	137	7.21	0	CAM	95	6.80
0	CAVITATION	137	4.89	0	TEMPERATURE	94	4.99
0	REQUIREMENT	136	4.55	0	NEED	94	4.71
0	RING	133	5.80	0	PARAMETER	93	5.21
0	TORQUE	132	5.09	0	SOURCE	93	5.21
0	LEVEL	131	4.71	0	DELIVERY	93	5.47

Figure 4. Characteristic Word Index Generated with $W_k^{(f)}$

r_k	$Term_k$	W_k	D_k	r_k	$Term_k$	W_k	D_k
10	PUMP	2806	1.44	0	BAR	132	4.99
0	SEAL	1760	4.15	0	TORQUE	131	5.09
0	PRESSURE	1199	1.99	0	TOOL	129	5.63
10	NOISE	1093	4.27	0	CONTAMINATION	128	8.80
10	HYDRAULIC	1035	1.89	0	LIFT	127	5.80
0	SPEED	849	2.40	0	SAPROPELITE	127	8.80
10	GEAR	816	3.40	0	PROBLEM	124	4.80
0	CIRCUIT	744	3.75	0	AUXILIARY	124	6.80
0	FLUID	686	1.77	0	PACK	120	6.21
0	VALVE	563	3.59	0	SAND	119	7.21
0	POWER	556	2.39	20	RETARDER	118	8.80
0	VANE	556	4.04	0	LEAKAGE	111	5.21
0	FLOW	492	2.69	0	DUTY	109	7.21
0	TURBINE	437	4.63	0	ABEX	109	7.21
0	FACE	410	4.89	0	DENISON	109	7.21
0	COUPL	354	4.99	0	LEVEL	109	4.71
10	VIBRATION	306	5.34	0	PERFORMANCE	108	3.59
0	ENGINE	304	5.63	0	TROUBLE	107	7.80
0	PROPULSION	275	6.47	0	CAM	107	6.80
0	OIL	275	4.21	0	CHARACTERISTIC	105	4.04
0	WATERJET	269	6.47	0	REQUIREMENT	102	4.55
10	DISPLACEMENT	246	3.75	0	ENERGY	101	5.09
0	SOLID	239	4.80	0	EFFICIENCY	101	4.04
0	VARIABLE	238	3.84	0	HITACHI	101	6.47
0	BEAR	222	5.99	0	CLOS	101	6.47
0	TRANSPORT	217	5.21	0	PROPERTIE	99	5.09
10	ROTARY	215	3.71	0	MINING	99	7.21
0	THRUST	213	6.80	0	PETROLEUM	98	7.80
0	WATER	204	4.21	0	BLADE	98	6.21
0	CENTRIFUGAL	201	4.55	0	INTERNAL	98	4.89
0	LIQUID	195	4.34	0	MODEL	98	5.34
0	RING	192	5.80	0	COOLANT	98	8.80
0	VORTEX	187	7.21	20	SERVO	97	6.21
0	PISTON	186	4.15	0	DPR	96	8.80
0	IMPELLER	181	4.71	0	SLURRY	96	7.21
0	AIR	179	4.80	0	BELLOW	94	8.80
0	STAGE	179	4.71	0	OPEN	94	6.80
0	DRIVE	170	3.67	20	REGULATION	93	6.47
0	MEASUREMENT	160	4.63	0	VSUPR	89	8.80
20	HYDROSTATIC	151	5.80	0	CHAMBER	89	5.80
0	BHRA	146	4.63	0	JET	89	5.09
0	PLATE	146	6.80	0	FILM	88	7.80
0	DEVICE	145	4.89	0	INSULAT	86	7.21
0	OUTPUT	144	5.09	0	BLOWER	86	6.80
0	CAVITATION	144	4.89	0	ANTI	85	6.80
0	RANGE	140	2.46	0	DELIVERY	83	5.47
20	TRANSMISSION	137	4.34	0	MEAN	82	4.89
0	SIDE	134	5.47	0	PASS	82	6.80
0	VERTICAL	134	5.99	0	STROKE	81	8.80
0	STANDARD	134	4.47	0	STORAGE	81	7.80

Figure 5. Characteristic Work Index Generated With $W_k^{(s)}$

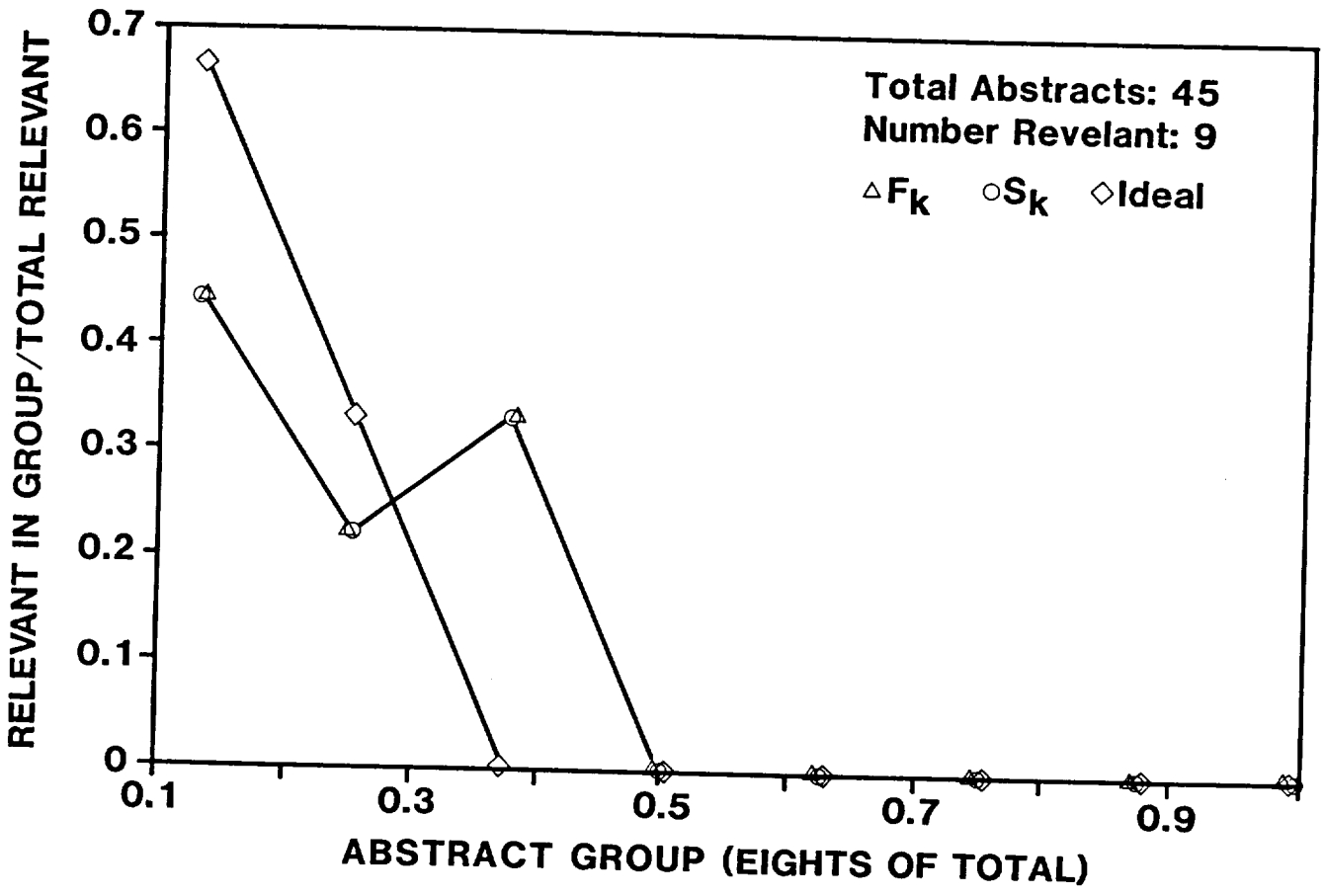


Figure 6. Citation Relevance

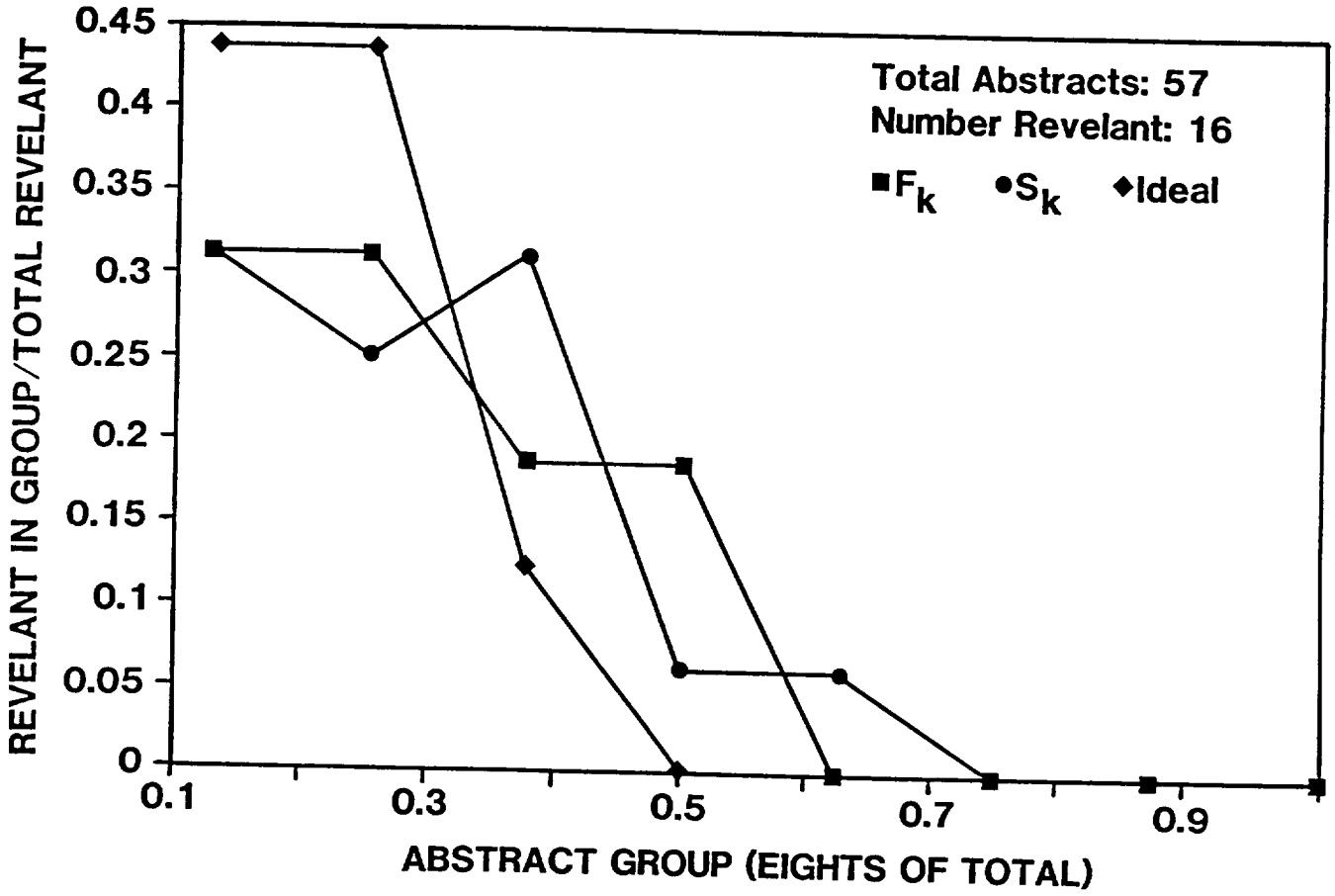


Figure 7. Citation Relevance

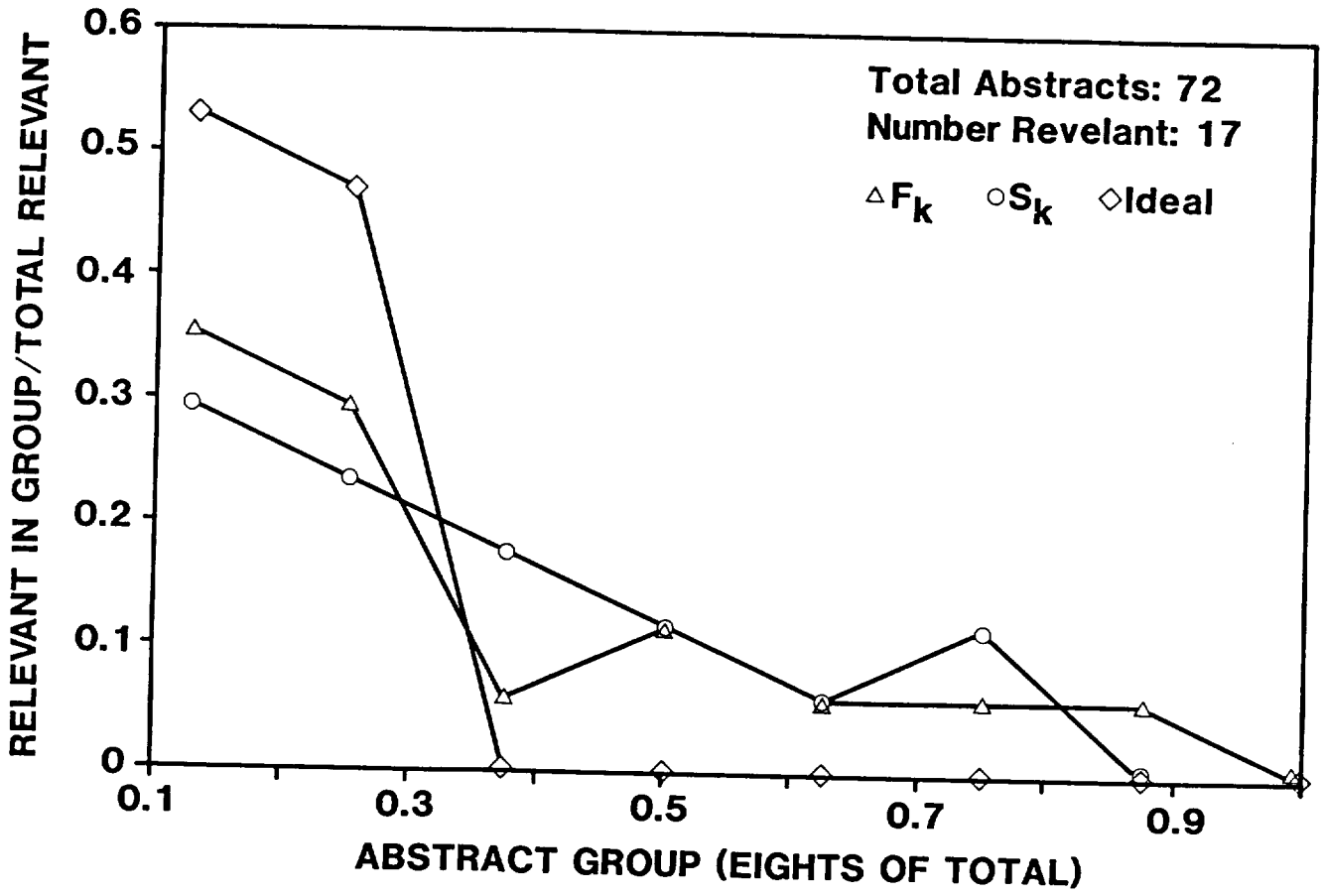


Figure 8. Citation Relevance

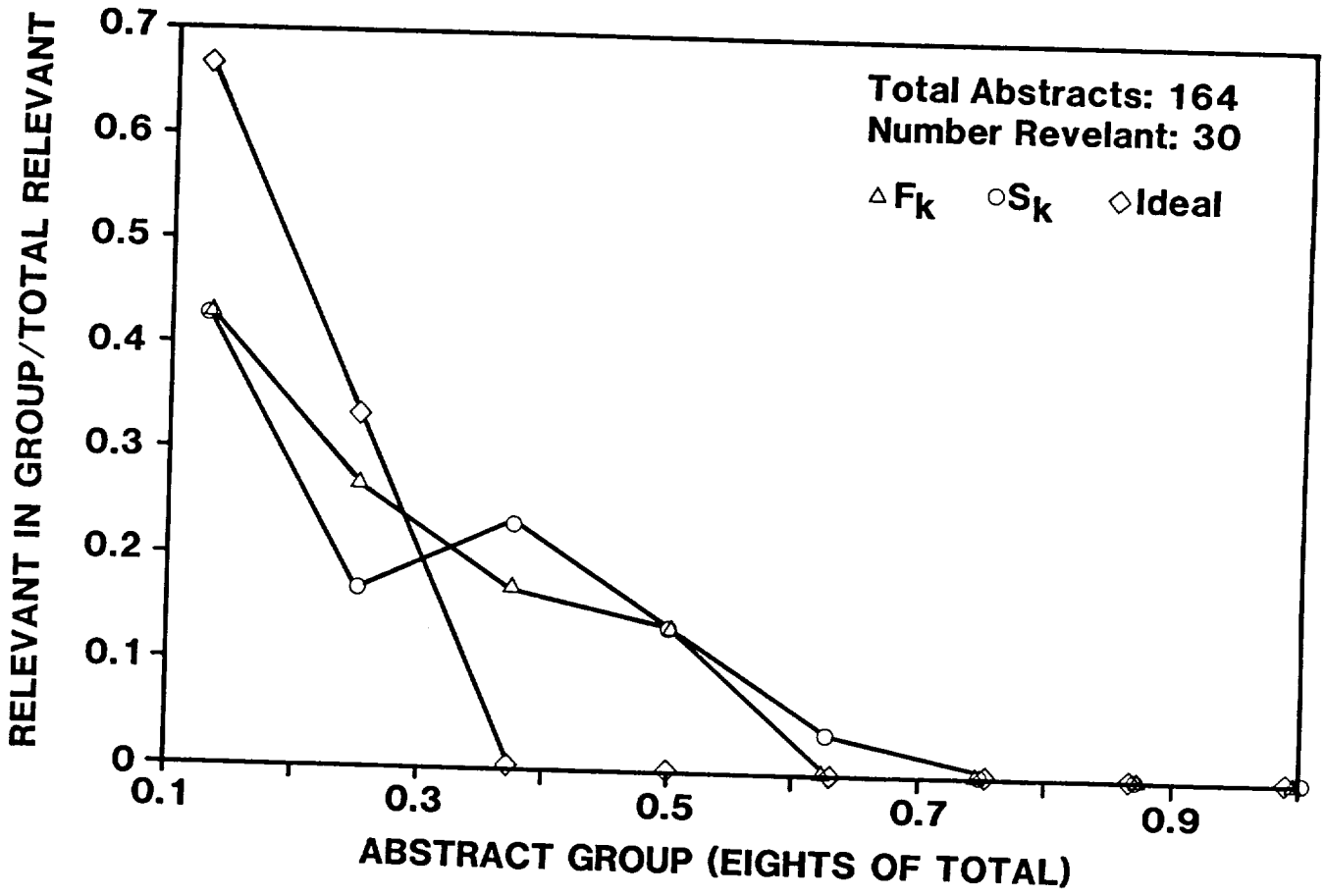


Figure 9. Citation Relevance

The results are generally positive with the algorithms providing significant concentration of citations when compared to a normal distribution. Furthermore, the results approach the ideal in a number of cases. The beneficial results are dependent upon significant computations which in turn require computing time. Some typical operational times for the two major algorithms--NABST and RANK--are shown in Figure 10. While not unduly prohibitive, the operational times--on an IBM AT computer--are substantial.

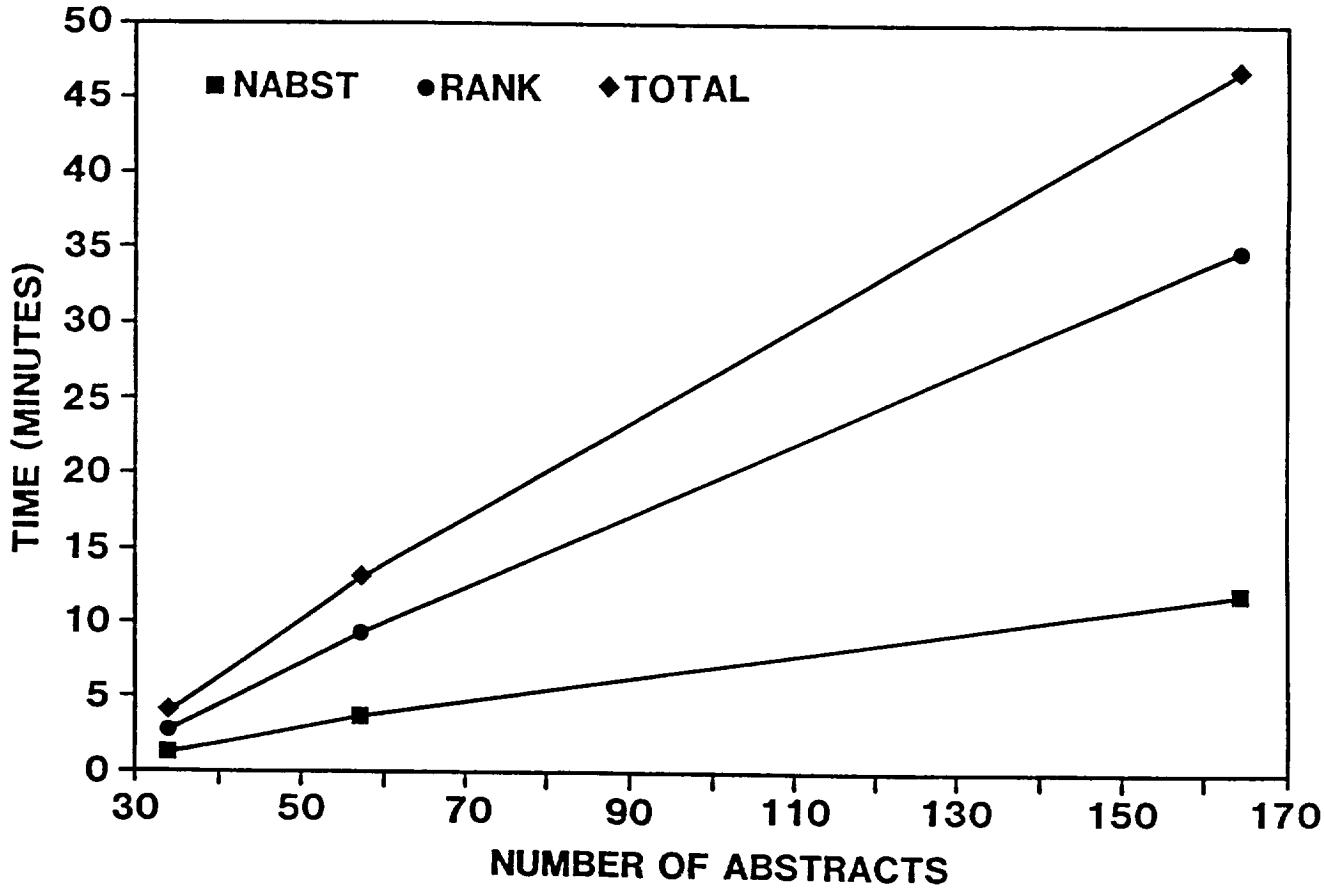


Figure 10. Computational Times for NABST and RANK on an IBM-AT Computer

1.5 Future Activities

A number of issues remain to be resolved. The overall effectiveness of the system must be verified in field tests. Preliminary results--while largely positive--are very limited in scope. The SORT-AID system is now being evaluated in a number of NASA Industrial Application Centers. These results will provide an objective and broadly-based assessment of the software system. The results--to date--are discussed in Section 2.

More sophisticated techniques could be used to generate the keyword index. These might prove to be more effective although order-of-magnitude improvements are unlikely. Reduction of computer run times is another area wherein overall effectiveness can be significantly improved. New generations of computer hardware will yield reduced operating times. Some initial experiments with the NABST algorithm recoded in the C language are very promising. Operating times were reduced by a factor of 3. If this holds for RANK as well, total operating times for both NABST and RANK would be on order of 5-6 sec/citation.

In summary, evaluations of the software systems will be continued. Reduced computer operational times will be emphasized in continuing developments.

2. USER EVALUATION OF SOFTWARE

2.1 Current Test Sites

The SORT-AID software system is currently installed at the Aerospace Research Applications Center (ARAC), the Southern Technology Assistance Center (STAC), the NASA Scientific and Technical Information Facility (STIF), and the University of Southern Mississippi. Additional sites will be activated during the coming months. User results to date along with a user's manual are discussed in the following sections.

2.2 Results to Date

ARAC has used the DEC 11/780 VAX version of SORT-AID extensively. Their results with NABST and REVIEW have been very positive, e.g., see [12]. ARAC has limited experience with RANK with mixed results. The current version of SORT-AID--PC-based--incorporates significant modifications to RANK. Recent results [13] indicate improved performance over earlier versions.

SORT-AID usage by both STAC and STIF has been limited. At this point, STAC's computer facilities are being utilized for a large database project. This has minimized the time available for SORT-AID evaluation. Preliminary results are promising but the software has not been used in the production environment. Completion of the database project along with acquisition of additional computer hardware should result in greater utilization of SORT-AID in the coming months.

2.3 User's Manual

A draft version of a SORT-AID user's manual [14] has now been developed. The manual is being distributed to the test sites for comment and review. A copy is included with this report. The user's manual includes a description of the citation review tool set, the relevance determination sub-system, an example SORT-AID session, installation of the software, the database information files, a description of the REVIEW commands and the stop list file.

2.4 Future Activities

The software system will be distributed to additional test sites during the coming year. Test results will continue to be analyzed and software modifications implemented as required.

3. ANCILLARY RESEARCH ACTIVITIES

In addition to the research and development documented directly in this report, a significant amount of ancillary and/or related research was carried out at the University of Southern Mississippi. This work has been documented in journal articles or in Master's theses and project reports. The various publications are listed in Table 1. Copies of all documents are available if desired.

TABLE 1

Publications Generated During
July 1, 1985 - June 30, 1986

Journal Articles

Huffman, G. David and Leigh, William. "The Well Equipped Searcher's Support Station", Microcomputers for Information Management, Vol. 3, pp. 59-68, 1986.

Leigh, William; Huffman, G. David; Paz, Noemi; and Vital, Dennis. "Development of a Technology Transfer Workstation", Proceedings of the Nineteenth Annual Hawaii International Conference on System Sciences, January 7-10, 1986, Honolulu, Hawaii.

Leigh, William; Burgess, Clifford; Huffman, G. David; and Paz, Noemi. "User Facilities for Engineering Support Stations", Proceedings of the Eighth Annual Conference for Computers and Industrial Engineering, March 19-21, 1986, Orlando, Florida. (Published as Journal of Computers and Industrial Engineering, Vol. 10, Supplement 1, 1986.)

Huffman, G. David and Leigh, William. "Semi-Automatic Determination of Abstract Relevance", Eleventh Annual Meeting and International Symposium--Technology Transfer Society, June 23-26, 1986, Indianapolis, Indiana.

Leigh, William, Huffman, G. David and Souder, R. "Aiding and Training Novice Computer Users Online with Executable Documentation", accepted by Journal of Educational Technology Systems, to appear summer, 1986.

Leigh, William, Huffman, G. David and Paz, Noemi. "A Mark-Up Language for the Presentation of Executable Documentation", Trends in Ergonomics/Human factors III, Elsevier Science Publishers, B.V. (North Holland), to appear 1986.

Huffman, G. David and Leigh, William. "A Technique for Presenting Computer System Instruction Online Using a Personal Computer", submitted to the Journal of Engineering Technology.

Leigh, William and Paz, Noemi. "Collecting and Transferring Searcher Expertise with a Personal Computer, Executable Documentation, and a Clear Box", submitted to Online Review, May, 1986.

Leigh, William and Paz, Noemi. "Using a Personal Computer to Provide Online and Executable Documentation for Searching Bibliographic Databases", submitted to Online, May, 1986.

TABLE ONE (CONTINUED)

Leigh, William and Paz, Noemi. "Using a Personal Computer to Provide Online and Executable Documentation for Simulation Models", submitted to Simulation, April, 1986.

Leigh, William, Paz, Noemi and Huffman, G. David. "SORT-AID with RANK: Tools for Automating the Determination of Citation Relevance", to be submitted to Information Processing and Management, June, 1986.

Huffman, G. David. "Semi-Automatic Determination of Citation Relevancy: A Preliminary Report", to be submitted to Information Processing and Management, July, 1986.

Master's Theses and Project Reports

Chiang, Pam and Lu, Hseuh-Ming. "Executable Documentation with Screen Division", joint M.S. project presented in fall, 1985.

McDonald, John. "A Simple Prolog Interpreter Written in XLISP", B.S. (honors) thesis presented in spring, 1986.

Chiang, Paul Wen-Sheng. "SEARCH-AID with the Mark-Up Language Interpreter", M.S. project presented in spring, 1986.

Yim, Roger Ki Song. "Automatic Generation of Online Documentation", M.S. thesis in progress.

Shih, Joan. "A LISP Interpreter with String-Processing Extensions", M.S. thesis in progress.

Vital, Dennis. "A Portable Bibliographic Database Searcher's Tool Set", M.S. thesis in progress.

Chiang, Pai-Ling. "PODS: A Polymorphic Online Documentation System for Transferring Bibliographic Database Search Expertise", M.S. thesis in progress.

REFERENCES

1. Huffman, G. David, Ulrich, J. M. and Bivins, R. G. "The Industrial Applications Study: A Mechanism for Technology Transfer." Technology Transfer Society, International Symposium Proceedings; Boston, Massachusetts; 1984, June.

2. Doszkocs, T. E. "Automatic Vocabulary Mapping in Online Searching." International Classification, 10(2): 78-83; 1983.

3. Marcus, R. S. "An Experimental Comparison of the Effectiveness of Computers and Humans as Search Intermediaries." Journal of American Society for Information Science, 34(6): 381-404; 1983.

4. Pollitt, A. S. "A Front-end System: An Expert System as an Online Search Intermediary." ASLIB Proceedings, 36(5): 229-234; 1984.

5. Fidel, R. "Towards Expert Systems for the Selection of Search Keys." Journal of American Society for Information Science, 37(1): 37-44; 1986.

6. Salton, G. "A Theory of Indexing." Society for Industrial and Applied Mathematics; Philadelphia, Pennsylvania; 1975.

7. Salton, G. Automatic Information Organization and Retrieval, McGraw-Hill Book Company; 1968.

8. Salton, G. Dynamic Information and Library Processing, Prentice-Hall, Inc.; 1975.

9. Salton, G. and McGill, M. J. Introduction to Modern Information Retrieval, McGraw-Hill Book Company; 1983.

10. van Rijsdergen, C. J. Information Retrieval, Butterworths, 1980.

11. Shannon, C. E. "A Mathematical Theory of Communication." Bell Systems Tech. J., 27: 379-423, 623-656; 1948.

12. Goehring, M. "Automating the Search Process--Phase I--SORT-AID Implementation and Characterization", Eleventh Annual Meeting and International Symposium--Technology Transfer Society, June 23-26, 1986, Indianapolis, Indiana.

13. Huffman, G. David. "Semi-Automatic Determination of Citation Relevancy: A Preliminary Report", to be submitted to Information Processing and Management, July, 1986.

14. Leigh, William; Paz, Noemi; Vital, Dennis; and Huffman, G. David. "SORT-AID: A Search Post-Processing Tool Set--User's Manual", University of Southern Mississippi, Hattiesburg, MS, July, 1986.